Available online at www.sciencedirect.com
ScienceDirect
Gene 381 (2006) 24–33
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Differences in the pattern of evolution in six physically linked genes of *Drosophila melanogaster*

Martina Žurovcová [a,*], Andrey Tatarenkov [b], Luděk Berec [a]

[a] *Institute of Entomology, Czech Academy of Sciences and Faculty of Biological Sciences, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czech Republic*
[b] *Department of Ecology and Evolutionary Biology, University of California, 321 Steinhaus Hall, Irvine, CA 982697-2525, U.S.A.*

Received 29 January 2006; received in revised form 2 May 2006; accepted 1 June 2006
Available online 28 June 2006
Received by A. Bernardi

Abstract

We describe DNA sequence polymorphisms at six loci (*Acp26Aa*, *Acp26Ab*, *Acp29AB*, *Idgf1*, *Idgf3* and *Ddc*), all on the second chromosome of *Drosophila melanogaster* in one natural European population. Previous studies considering these loci separately showed that some of them were affected by natural selection. However, demographic processes or population admixture can produce footprints similar to natural selection. Simultaneous consideration of several genes may help to discern between selective and demography/admixture scenarios because the latter are expected to affect a majority of loci in a similar manner. Such an effect is not necessarily uniform among genes, but can be modified by rates of recombination and substitution. Since different evolutionary forces shaped the variation of the studied genes, our aim is to examine if their physical linkage could have affected the observed pattern. Fisher's conservative test of linkage disequilibrium is not significant. Lewontin's sign test pointed to linkage disequilibrium both within and between loci levels, though, none of the loci exhibits haplotype structure. Coupled with other results, the possibility of demography being the exclusive explanation for the observed variability is ruled out.

Keywords: DNA polymorphism; Demography; Linkage disequilibrium; Molecular evolution

1. Introduction

Until now a number of studies dealing with multiple-loci screening from natural populations of *Drosophila melanogaster* have been conducted. Some of those studies were focused on general patterns of the demographic and selection history of natural populations (Glinka et al., 2003; Baudry et al., 2004; Orengo and Aguadé, 2004; Haddrill et al., 2005). Others were dedicated to more specific questions such as the relationship among linkage disequilibrium, recombination and gene conversion (Langley et al., 2000; Andolfatto and Wall, 2003), evolution of the particular group of proteins (Begun et al., 2000) or detecting the footprint of a selective sweep (Sáez et al., 2003). Studies of individual genes are bound to remain important if one is interested in intricacies of natural selection and its connection with the functional importance of the genes (Lazzaro and Clark, 2003).

However, if genes are considered individually, the effects of natural selection can sometimes be difficult to disentangle from the effects of certain demographic processes or population admixture, which leave similar changes in nucleotide variation (Przeworski, 2002). Since demography should impact a majority of genes, while natural selection is more idiosyncratic, the simultaneous consideration of several genes is generally considered as a way to differentiate between these factors (Andolfatto and Przeworski, 2000; Glinka et al., 2003). The interaction of demographic processes with natural selection makes nucleotide variation even more complex. Accounting for linkage disequilibrium is a key area where a more detailed analysis can bring new insights (Nordborg and Tavaré, 2002). However, interaction of linkage disequilibrium and selection remains a contentious issue. Nurminsky et al. (2001)

Abbreviations: *Acp26Aa*, Accessory gland-specific peptide 26Aa; *Acp26Ab*, Accessory gland-specific peptide 26Ab; *Acp29AB*, Accessory gland-specific peptide29AB; *Idgf1*, Imaginal disc growth factor 1; *Idgf3*, Imaginal disc growth factor 3; *Ddc*, Dopa decarboxylase; *Adh5′*, 5′ end region of Alcohol dehydrogenase; PCR, polymerase chain reaction; bp, a base pair; kbp, a thousand base pairs; FET, Fisher's exact test; LD, linkage disequilibrium.

* Corresponding author. Tel.: +42 38 777 5260; fax: +42 38 5310354.
E-mail address: martina@entu.cas.cz (M. Žurovcová).

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.06.011

reported selection affecting linked loci over very large scale (two cytological divisions) and Sáez et al. (2003) recently detected recent selective sweep over 41–54 kbp, but Andolfatto et al. (1999) showed that a selection effect can decline quickly even within very close proximity of a selected region.

Recently, several studies suggested that in non-African natural populations of *D. melanogaster*, the effect of demography is quite considerable (Andolfatto and Przeworski, 2000; Glinka et al., 2003; Baudry et al., 2004; Caracristi and Schlötterer, 2003; Ometto et al., 2005; and others). Although these studies made a huge contribution with respect to the elucidation of the history of both African and European populations, timing of the presumed bottleneck affecting the non-African populations is still unclear. Most of the analyses (Glinka et al., 2003; Ometto et al., 2005; Haddrill et al., 2005; Thornton and Andolfatto, 2006) were also based on a single sample from Netherlands, which means that the results may be specific to this particular population. Orengo and Aguadé (2004) actually showed that unlike the Netherlands population, the observed pattern in a population from Spain cannot be explained by a bottleneck scenario alone, even when using the same type of data (non-coding loci on X-chromosome).

It was our aim to revise the previously published information and to extend the analyses in these previously unexplored directions. Our analysis uses genes from the left arm of the second chromosome, which have been recently screened for genetic variability: *Acp26Aa, Acp26Ab* (Aguadé, 1998), *Acp29AB* (Aguadé, 1999), *Idgf1, Idgf3* (Žurovcová and Ayala, 2002), and *Ddc* (Tatarenkov and Ayala, unpublished). Studied from the same individuals of a single population (Montblanc, Spain), these loci provide a good opportunity for multiple locus comparisons. Some loci are closely localized; coding regions of *Acp26Aa* and *Acp26Ab* are separated just by ~200 bp, and the loci of another pair, *Idgf1* and *Idgf3*, are 4 kbp apart. On the other hand, 11 cytological bands (from 26A to 37D) containing about 10 Mbp separate the two most distant genes *Acp26Aa* and *Ddc*. *Ddc* is also one of the genes for which a considerable linkage disequilibrium was described both at the larger scale of 65 kbp (Aquadro et al., 1992) as well as at the shorter scale (Tatarenkov and Ayala, unpublished) within the gene.

The genes have different functions and represent different evolutionary models. *Acp26Aa* and *Acp26Ab* encode for sex-related peptide hormones (male accessory gland proteins). These loci are tightly linked in a region with a high recombination rate ($c_{lab}=4.402\times10^{-8}$ rec/bp/generation). Another locus, *Acp29AB*, belongs to the same functional family of male accessory gland proteins but does not show similarity to the other *Acps* ($c_{lab}=4.286\times10^{-8}$ rec/bp/generation). Development-related genes, *Idgf1* and *Idgf3* (imaginal disc growth factors), encode for proteins with growth promoting activity. They are members of a multigene family and localized in a region with a medium recombination rate ($c_{lab}=1.611\times10^{-8}$ rec/bp/generation). *Ddc* (dopa decarboxylase), whose protein product is involved in dopamine biosynthesis (neurotransmitter and cuticle formation), also belongs to a multigene family and is located in a region with the lower recombination rate ($c_{lab}=0.774\times10^{-8}$ rec/bp/generation).

There are two goals in this study. By examining the variability at the previously screened autosomal genes, we aimed to get additional insights by focusing on the linkage disequilibrium at the intralocus and interloci levels, with particular attention to the recombination rate issue. This goal is intertwined with the second task, which was to test the bottleneck hypothesis by combining the approach of Orengo and Aguadé (2004), whose sample also originated from Spain, and Lazzaro and Clark (2003), whose analysis of non-African population was based on autosomal loci.

2. Materials and methods

2.1. Data acquisition

D. melanogaster data are from lines isogenic for the 2nd chromosome, derived from flies collected on Montblanc (Tarragona, Spain; Aguadé, 1998, 1999).

Data for *Acp26Aa* and *Acp26Ab* (both $n=10$) were from Aguadé (1998), *Acp29* ($n=12$) from Aguadé (1999), *Idgf1* and *Idgf3* (both $n=20$) from Žurovcová and Ayala (2002) and *Ddc* ($n=15$) from Tatarenkov and Ayala (unpublished). The sequences were either taken from GenBank (all *Acp* genes; alignments made de novo) or directly provided by the authors (*Idgfs*, *Ddc*; alignments as in the original publications). A summary of the screened lines in different studies is given in Supplementary Table 1.

Sequences were edited and aligned both manually and with the assistance of the EditSeq and MegAlign programs (Lasergene package of programs from DNASTAR, Inc.). Physical locations of the individual genes were obtained from release 4.0 of the *D. melanogaster* genome available at FlyBase. Distance between the loci was estimated from their cytogenetical position and the cumulative parameter provided by the Recomb-Rate program (Comeron et al., 1999). The original population was cytogenetically inspected for the common inversions, which were found to be very low as expected for a European population (M. Aguadé, personal communication). Additionally, we inspected lines sequenced for *Idgfs* and *Ddc* for the presence of the *In(2L)t* inversion by the PCR-based method developed by Andolfatto et al. (1999).

2.2. Statistical analysis

Basic population genetics parameters were obtained using DnaSP version 4.00 (Rozas et al., 2003). Nucleotide diversity π, the average number of nucleotide differences per site between two sequences, was estimated according to Nei (1987, Eqs. 10.5 or 10.6). The population mutation parameter θ (per site) was based on the number of segregating sites S (Watterson, 1975, Eq. 1.4a, but on a base pair basis; Nei, 1987, Eq. 10.3).

Most of the neutrality tests were conducted using DnaSP version 4.00 (Rozas et al., 2003). DNA Slider (McDonald, 1998) was used to explore the heterogeneity in the polymorphism-to-divergence ratio along the individual loci. We used the Combo software kindly provided by Lewontin to perform the Lewontin's test of linkage disequilibrium (Lewontin, 1995). The presence of possible gene conversion events was tested by Geneconv (Sawyer, 1999), and LDhat (McVean, 2004) was used for some parts of the recombination rate analysis.